Blogs and News Sources Coverage in Altmetrics Data Providers: A Comparative Analysis by Country, Language, and Subject

José Luis Ortega

Institute for Advanced Social Studies (IESA-CSIC), Córdoba, Spain, jortega@iesa.csic.es

Joint Research Unit Knowledge Transfer and Innovation, (UCO-CSIC), Córdoba, Spain

Abstract

The main objective of this study is a detailed analysis of the coverage of blogs and news by three of the most important altmetrics data providers (Altmetric.com, PlumX, and Crossref Event Data). Concretely, the study looks for differences in blogs and news coverage, according to three criteria: country, language, and subject, with a view to detecting biases that influence altmetrics impact. More than 100,000 random publications from Crossref were searched in all three providers. The link, title, and source of the events that mention each document were retrieved. Each source was classified according to the three criteria. Results show that over 65% of blogs and news come from English-speaking countries and over 75% are written in English. In terms of subject matter, General-interest news outlets (> 50% in PlumX and Altmetric.com) and Social Sciences and Humanities blogs (> 20%) prevail. Altmetric.com is the most geographically and linguistically heterogeneous service, with the best coverage of blogs; PlumX collects more news media, especially local-interest newspapers from the United States; and Crossref Event Data is the platform that brings together most English-speaking sources.

Keywords

Altmetrics, Data providers, Blogs, News, Coverage

Introduction

Altmetrics measure research impact, complement the information of other quantitative or qualitative indexes (Holmberg, 2015), and provide valuable evidence of how a scientific study is currently perceived by different web environments. The mention of scholarly articles in news, blogs, social networks, or patents provides a window on how a research product influences public opinion (social networks), innovation (patents), mass media (news and blogs), or government (reports and policy documents). In this manner, altmetrics emerge as possible indicators of knowledge transfer because they enable the quantification of the connections of scholarly activity with other spheres of knowledge, thereby contextualizing the role of science in society.

Until now, altmetrics data providers have been the easiest and most accessible way to obtain these indicators. These services monitor social networks (Twitter, Facebook, etc.), news outlets, blogs, and many web sites, searching for events that mention scientific results. This harvesting process captures the impact of academic results in a great variety of fields. Depending on the metric being analyzed, these platforms use different sources. Many of them are events that occur in only one web space, such as Twitter tweets, Mendeley's readers, or Facebook likes.

However, many other metrics come from a heterogeneous range of sources such as blogs and news. In these cases, altmetrics providers commonly manage an internal list of blogs and news media where they search for mentions of scholarly outputs. In addition to these lists, these services contract specialist providers that supply data on academic mentions from a broader range of media. Altmetric.com uses Moreover.com, a clipping service, to obtain news mentions, while PlumX counts on the ACI Scholarly Blog Index to extract mentions from blogs.

Irrespective of the way in which each service obtains data, information about the coverage of blogs and news, in particular, is not easily accessible. Altmetric.com only publishes a list of images of mainstream media outlets (https://www.altmetric.com/about-our-data/our-sources/news/), where it is impossible to link each medium with the event in question, or to discover the correct name. Moreover, there is no information about blogs. Worse still, PlumX provides no details about the list of blogs and news it covers. Crossref Event Data (CED), on the other hand, makes it possible to extract information about its sources, although they are merely web domains with no name, origin, or type.

Faced with the absence of reliable facts about the coverage of blogs and news media in altmetrics providers, this study attempts to shed light on the coverage of the three most important services as a way to compare them and to detect biases that could distort the impact perceived through these platforms.

Literature Review

One of the most important research lines in altmetrics is the study of data provider coverage. The validity and reliability of altmetrics research relies on the functionality of these tools, since any bias or error can seriously influence results. Although Adie and Roe (2013) were the first to describe the features of Altmetric.com, it was not until Robinson-García et al. (2014) explored the coverage of this service that the selection of external sources was deemed to require further research. Hassan et al. (2017) indicated a rapid increase in the presence of Altmetric.com data in the Scopus database. PlumX has deserved less attention. Champieux (2015) was the first to provide a detailed outline of the tool, and Jobmann et al. (2014) used this aggregator for altmetrics studies for the first time to compare the coverage of three data aggregators. Torres-Salinas et al. (2017) tested the utility of PlumX for books and found that the usage statistics of this service offered original information about the success of monographs. CED has been examined far less and more recently; however, we would highlight works by Zahedi and Costas (2018) and Ortega (2018) that employ these data to compare the coverage of altmetrics data services.

The existence of different services supplying altmetrics data has given rise to several studies that have made a comparative analysis of the coverage of these platforms. The first attempt was by Jobmann et al. (2014), who compared the production of the Leibniz Association across four platforms (PlumX, Altmetric.com, ImpactStory, and Mendeley). Their results showed that PlumX obtained more Facebook posts, whereas Altmetric.com received more mentions on Twitter. Zahedi et al. (2015) investigated the consistency of Altmetric.com, Mendeley, and Lagotto, finding that Mendeley had the highest coverage (69%), followed by Lagotto (68%) and Altmetric.com (23%). Peters et al. (2016) studied the coverage of data sets in PlumX, ImpactStory, and Altmetric.com. Of the three altmetrics tools, they found PlumX to have the best coverage. More recently, Meschede and Siebenlist (2018) correlated the metrics of Altmetric.com and PlumX, observing that Mendeley readers and Wikipedia citations showed high correlations, while News was the count that correlated less. Ortega (2018) compared the

counts of more than 50,000 articles in PlumX, Altmetric.com, and CED. The results showed that while Altmetric.com captures more tweets, blogs, and news mentions, PlumX collects more Mendeley readers, and CED has a special coverage of Wikipedia citations. In the same vein, Zahedi and Costas (2018) performed the most complete analysis, checking five altmetrics services. They also found serious imbalances in the distribution of counts by provider. Torres-Salinas et al. (2018) compared the coverage of books by Altmetric.com and PlumX, concluding that none of these tools provide a complete picture of the social attention generated by books and that they are complementary rather than comparable.

Few papers have explored the particular role of blogs and news in altmetrics impact and their coverage by data providers. Shema et al. (2014) studied the mention of research articles in ResearchBlogging.org and their citations in Web of Science. They found a significant association, suggesting that blog citations could be a suitable altmetrics indicator. Bornmann (2014) explored a set of articles from Altmetric.com and observed a limited number of documents mentioned in blogs (16%) and news (13%). However, Fraumann et al. (2015) were the first to explore the content of the blogs and news covered by Altmetric.com, observing considerable bias toward English-speaking countries. Finally, Ortega (2019a) analyzed blogs and news coverage in three data providers, observing very low overlap and concluding that several providers are required to ensure a reliable altmetrics study. However, no study has compared the country, language, and subject distribution of blogs and news by different providers.

Objectives

The principal objective of this study is to describe the distribution of blogs and news in the three major altmetrics providers: Altmetric.com, PlumX, and CED, according to the country, language, and subject matter of each medium. These distributions reveal differences and similarities between providers and possible biases that could affect the altmetrics impact of publications. Three research questions were formulated:

- What are the differences/similarities between altmetrics providers, according to the geographical distribution of blogs and news?
- What is the percentage of blogs and news by language? And how does this percentage change, according to the different providers?
- Which subjects are more frequent in blogs and news? And how are they distributed across the three altmetrics providers?

Methods

Altmetrics Providers

PlumX: PlumX (plu.mx/plum/g/samples) is a provider of alternative metrics created in 2012 by Andrea Michalek and Michael Buschman from Plum Analytics. This product targets the institutional market, offering altmetrics counts of publications for private institutions. PlumX is the aggregator that offers the most metrics, including citation and usage metrics (i.e., Views and Downloads) and is the largest altmetrics aggregator (Plum Analytics, 2019), covering more than 52.6 million artifacts. In 2017, Plum Analytics was acquired by Elsevier (www.elsevier.com) and now tracks the online presence of any article indexed in the Scopus database (Elsevier, 2017).

Altmetric.com: The first altmetrics provider was set up in 2011 by Euan Adie, with the support of Digital Science (www.altmetric.com). Unlike PlumX, Altmetric.com focuses on the publishing world, signing agreements with publisher houses to monitor the altmetrics impact of their

publications. This information is accessible through a public Application Programming Interface (API). Today, Altmetric.com tracks the social impact of close to nine million research papers (Altmetric.com, 2019a).

Crossref Event Data: CED is the newest service. Created in 2016, it is still in beta format (www.crossref.org/services/event-data). Unlike the others, CED is not a commercial site and provides free access to data through a public API. Another important difference is that it does not provide metrics, but only displays information about each altmetrics event linked to a Digital Object Identifier (DOI). For instance, it shows the information about the mention of an article on Twitter (date, user, tweet, etc.), but it does not show a count of the number of tweets. For this reason, CED data would require processing in order to be comparable with the other services.

Sources: Blogs and News

Basically, two different approaches are used to compile blogs and news mentions. The first draws up a list of venues where events are tracked, and the second selects an external specialized service that provides this information. Both practices have been equally used by PlumX and Altmetric.com to collect one metric or another. CED, however, simply crawls specific domains to extract mentions, with no distinction between blogs and news. Because these providers do not establish a clear definition of blog and news outlet, some sources are equally classified as blogs and news at different times. This error is not common, only occurring in 3% of the events in PlumX and Altmetric.com (Ortega, 2019a).

Blogs

PlumX includes blogs from 2015 (Parkhill, 2015), extracting mentions from an internal list. In 2016, Plum Analytics extended the list of 4,000 sources to more than 10,000 blogs provided by the ACI Scholarly Blog Index (Parkhill, 2016). However, this source is no longer active, neither are many of its links (Ortega, 2019b). Another serious problem is that some sources considered as blogs are in fact research journal websites (Hindawi and Omic Publishing Group), a proportion estimated at around 23.7% (Ortega, 2019a). From the beginning, Altmetric.com has curated a manual list of roughly 14,000 blogs (Altmetric.com, 2019b), although it does not publish the list of sources. CED makes no clear distinction between blogs and news. It defines three categories for grouping web domains: wordpressdotcom, web, and newsfeeds (Crossref, 2019). The first group belongs to sites hosted by WordPress, though they are not strictly blogs. The second is only websites, which could include blogs or not. In addition, the reddit-links category includes links from Reddit that connect to external sources such as blogs and news.

News

The agreement between PlumX and Elsevier entailed the use of Newsflo (an Elsevier company) as a news data provider. With this agreement, PlumX covers more than 55,000 different news sources (Allen, 2017). In the manner of PlumX, Altmetric.com used Moreover.com, a news aggregator. This collaboration allowed Altmetric.com to cover more than 80,000 news outlets, in addition to the initial list of 1,300 (Williams, 2015). However, this company was acquired by Lexis-Nexis and the collaboration ended, leaving 19% of the links inactive (Ortega, 2019b). Altmetric.com currently manages a list of 2,900 news outlets (Altmetric.com, 2019). The name of the media is publicly available on the site. CED uses the *newsfeed* category to group news sources, but a manual inspection revealed that media and blogs are classified equally as *web* or *newsfeed*, and sometimes in both categories. Therefore, the distinction between blog and news

is based on the matching with other data providers. In the event of a non-match, mentions are then classified manually.

Data Extraction

This study aims to compare the coverage of blogs and news mentions by the three major altmetrics providers. A random sample of 100,529 DOIs from CED were extracted to detect the number of publications covered by these aggregators. These publications were obtained from CED API as journal articles, published from 2012 on (https://api.crossref.org/works?sample=100&filter=type:journal-article,from-pub-date:2012-01-01). The year 2012 was chosen because it provides a sufficiently broad time window to capture the impact of the sample on blogs and news.

Next, this list was searched in all three data providers. In the case of Almetric.com, the Altmetric ID was obtained from the Altmetric API (api.altmetric.com/v1/doi/) and was used to extract data about blogs and news directly from the website (www.altmetric.com/details/), because the API only shows counts but not the links or content of these mentions. In the case of PlumX, DOIs were searched on the PlumX website (plu.mx/plum/a/?doi=). Finally, information about CED was extracted from the API (query.eventdata.crossref.org/events?filter=obj-id:). In all three cases, several SQL scripts were written to scrape the data from websites and APIs. This process was performed in the second half of August 2018.

Classification Criteria

Because no single altmetrics provider classifies its sources, this study set up three criteria to group them, in order to make a fair comparison between services:

Country: This criterion groups sources according to the country in which the person responsible is located. The purpose is to show the geographic distribution of blogs and media and to detect biases in the coverage of media by country. A manual inspection identified the country of residence by exploring sections such as "About us" and "Contact". In the case of several authors based in different countries (i.e., 2Physics, www.2physics.com), the label "International" was applied. News media with different country editions (i.e., Huffington Post, huffingtonpost.ca, huffingtonpost.fr) were considered different media and classified in each country.

Language: The language of the website was used to group sources. This criterion was established to observe the influence of language on the mention of research papers and to point out the extent to which altmetrics providers cover non-English-speaking media. Language information was extracted from meta tags (i.e., <html lang="en">, <meta http-equiv="Content-Language" content="en">). Where pages lacked this information, a manual inspection was carried out. When several languages were present in the source, the most common was selected.

Subject: This last criterion attempts to group sources according to content. The aim is to observe the subject matter distribution of these websites and to detect thematic biases among providers. Titles and keywords from the metadata were used to assign the webpage to a category. In the event of doubt, a manual inspection was made. All Science Journal Classification Codes (ASJC) (Scopus, 2019) were used to group the content of the sources by category. This system was selected because it is multidisciplinary, is widely used (Scopus, Publons, QS, etc.), and enables the comparison of sources with publications. However, this classification scheme was designed for scholarly journals and not for blogs and news media. In this case, some adaptions were made and new categories were added:

- General-interest: This category was created to group news media covering news from any category, such as politics, economics, science, and sports, and addressed to a general audience (i.e., *Washington Post*, *The Guardian*).
- Local-interest: This group was assigned news media with a clear local component, namely, content addressed to a local audience (i.e., *ABC News 15 Arizona*, *Boise State Public Radio*).
- Science and Technology: This group brings together news media and blogs specializing in scientific topics, including news about physics, biology, health, engineering, social sciences, and history (i.e., *Discover Magazine*, *e! Science News*).
- Entertainment: Finally, this group was created to collect blogs and media that discuss Fitness, Beauty, Cooking, Games, etc. (i.e., *Cosmopolitan*, *Organic Authority*).

Results

Overall, Altmetric.com is the provider that gathers most sources in the sample (3,856), followed closely by PlumX (3,255), and CED (1,263). However, the number of news and blogs in each service is rather different. Altmetric.com presents many more blogs (2,582) than news outlets (1,435), while PlumX includes fewer blog sources (860) and more news sites (2,559). CED has a very low number of news (310) in relation to blogs (1019).

Country

The first comparison criterion is the coverage of sources by country, which reveals whether the selected sources are grouped in a few countries or spread over several. More than 95% of the sources in Altmetric.com and PlumX were assigned to a single country; the remaining sites do not provide location information. In the case of CED, only 79.8% of the blogs were classified. This low proportion is because CED includes many inactive blogs that were created as an exercise in academic courses (i.e., eportfoliopsy1100.wordpress.com). These sites do not include geographical information.

Figure 1 and Table 1 display the distribution of news outlets by country. Figure 1 ranks the countries by the percentage of news media in Altmetric.com. Table 1 shows the first ten countries in each provider by number of news sources. Four of the ten countries in each rank are English-speaking, contributing 87.7% of the sources in CED, 86.7% in PlumX, and 67.1% in Altmetric.com. These added percentages demonstrate a clear bias toward these countries.
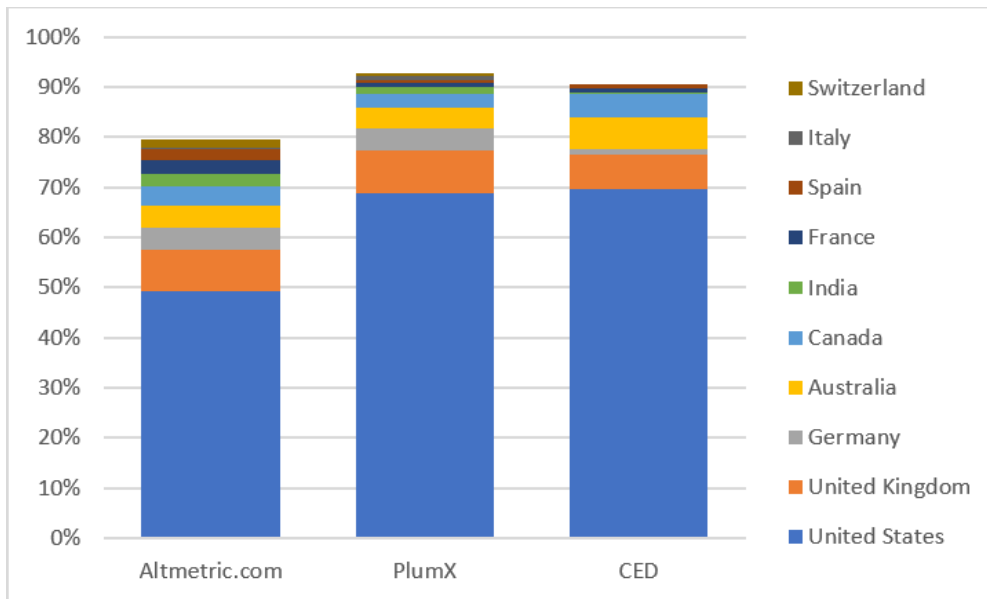
Figure 1. Percentage of news outlets by country in the three major providers limited to ten first countries.

Altmetric.com			PlumX			CED		
Country	Sources	%	Country	Sources	%	Country	Sources	%
United States	706	49.2%	United States	1762	68.9%	United States	216	69.7%
United Kingdom	121	8.4%	United Kingdom	174	6.8%	United Kingdom	32	10.3%
Germany	63	4.4%	Australia	162	6.3%	Canada	14	4.5%
Australia	61	4.3%	Canada	71	2.8%	Australia	9	2.9%
Canada	54	3.8%	South Africa	41	1.6%	International	7	2.3%
India	39	2.7%	India	33	1.3%	Italy	3	1.0%
France	38	2.6%	Germany	29	1.1%	France	3	1.0%
Spain	31	2.2%	France	22	0.9%	Germany	3	1.0%
Italy	29	2.0%	Switzerland	14	0.5%	Spain	2	0.6%
Switzerland	20	1.4%	Spain	12	0.5%	Russia	2	0.6%
Not Assigned	25	1.7%	Not Assigned	40	1.6%	Not Assigned	5	1.6%
TOTAL	1,435	100%	TOTAL	2,559	100%	TOTAL	310	100%

Table 1. Number and percentage of news outlets by country in the three major providers. First ten countries.

With regard to the specific position of each country, the United States contributes more sources in all three aggregators, from 49.2% in Altmetric.com to 69.7% in CED. Next is the United Kingdom with 10.3% in CED, 8.4% in Altmetric.com, and 6.8% in PlumX. It is worth mentioning the important presence of Indian media (2.7%) in Altmetric.com, the highlighted coverage of news outlets from South Africa (1.6%) in PlumX, and the significant presence of international media (2.3%) in CED.

If the percentage of media of the first ten countries is added, it would indicate the degree of concentration/spread of the media covered. Thus, Altmetric.com is the site that gathers news media from a wider variety of countries (79.9%), while PlumX (91.8%) and CED (90.1%) concentrate most of their sources in the first ten countries. This fact is observed in Figure 1, in which the total size of the bars expresses the indicator of concentration/spread.

Figure 2 and Table 2 display the distribution of blogs by country. Figure 2 ranks countries by the percentage of blogs in Altmetric.com. Table 2 shows the first ten countries in each provider by number of blog sources. In terms of news, the countries that contribute more blogs are English-speaking countries: 80.5% in PlumX, 77.4% in Altmetric.com, and 69.2% in CED.
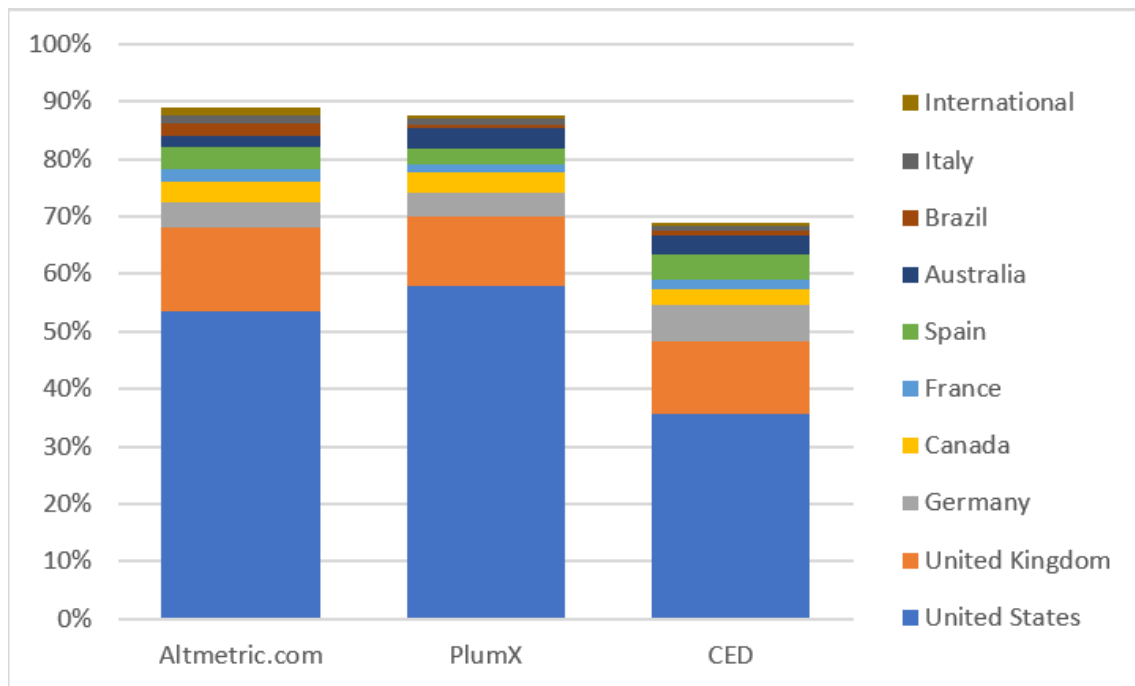

Figure 2. Percentage of blogs by country in the three major providers limited to ten first countries.

Altmetric.com			PlumX			CED		
Country	Sources	%	Country	Sources	%	Country	Sources	%
United States	1,383	53.6%	United States	503	57.9%	United States	363	35.6%
United Kingdom	376	14.6%	United Kingdom	101	12.1%	United Kingdom	128	12.6%
Germany	96	3.7%	Australia	31	3.6%	Australia	34	3.3%
Canada	93	3.6%	Canada	30	3.5%	Spain	29	2.8%
France	60	2.3%	Spain	21	2.4%	Canada	28	2.7%
Spain	53	2.1%	International	15	1.7%	Germany	22	2.2%
Australia	53	2.1%	Germany	13	1.5%	Brazil	20	2.0%
Brazil	44	1.7%	France	12	1.4%	France	17	1.7%
Italy	35	1.4%	Italy	11	1.3%	China	14	1.4%
International	24	0.9%	Netherlands	11	1.3%	India	12	1.2%
Not Assigned	121	4.7%	Not Assigned	29	3.3%	Not Assigned	206	20.2%
TOTAL	2,582	100%	TOTAL	869	100%	TOTAL	1,019	100%

Table 2. Number and percentage of blogs by country in the three major providers. First ten countries.

The United States is again the country that features more blogs in all three providers, although blogs are less prominent than news. Thus, PlumX collects 57.9% of blogs from the United States, Altmeric.com 53.6%, and CED 35.6%. Conversely, the United Kingdom increases its presence with 12.6% of blog sources in CED, 14.6% in Altmetric.com, and 12.1% in PlumX. The remaining

countries follow a similar distribution in all three aggregators, aside from the special coverage of German blogs (3.7%) in Altmetric.com.

According to the concentration/spread of sources, PlumX (86.5%) and Altmetric.com (87.9%) show a high concentration of sources in the first ten countries, while CED displays a higher spread (67.5%). This low value in CED is caused by the high number of unassigned blogs (20.2%).

Language

The second criterion is the proportion of sources according to different languages. This information complements country distribution and reveals the degree of dissemination and the audiences covered by altmetrics providers. More than 98% of the sources were assigned to a language, in the case of news. As for blogs, 7.1% in CED and 5.6% in Altmetric.com cannot be assigned to a language, the reason being that these pages were not active at the time of the study.

Figure 3 and Table 3 depict the distribution of news media by language. Figure 3 ranks the languages by the percentage of news in Altmetric.com. Table 3 shows the first ten languages in each provider by number of news sources. English-speaking news media are the most common sources in the three providers, ranging from 97.7% in CED to 76.7% in Altmetric.com. The remaining languages show low percentages and different proportions according to provider. Thus, in Altmetric.com, German is the second language, with 5.5%, and Spanish the third, with 3.6%. In PlumX, French is the second language, with just 0.9%, and in CED, Italian is the second, with 0.6%.
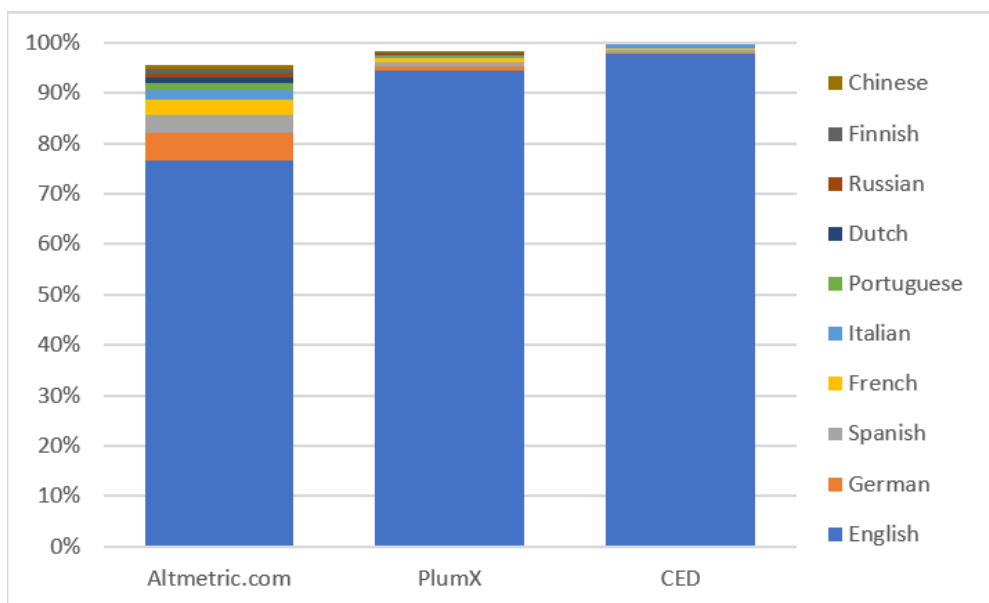

Figure 3. Percentage of news outlets by language in the three major providers limited to the first ten.

Altmetric.com			PlumX			CED		
Language	Sources	%	Language	Sources	%	Language	Sources	%
English	1,101	76.7%	English	2,420	94.6%	English	303	97.7%
German	79	5.5%	French	23	0.9%	Italian	2	0.6%
Spanish	51	3.6%	German	21	0.8%	Spanish	2	0.6%
French	43	3.0%	Spanish	18	0.7%	Norwegian	1	0.3%

Language			Language			Language		
Italian	28	2.0%	Italian	8	0.3%	French	1	0.3%
Portuguese	18	1.3%	Chinese	7	0.3%	German	1	0.3%
Dutch	17	1.2%	Russian	6	0.2%			
Russian	12	0.8%	Portuguese	4	0.2%			
Finnish	11	0.8%	Dutch	4	0.2%			
Chinese	10	0.7%	Finnish	4	0.2%			
Not Assigned	16	1.1%	Not Assigned	27	1.1%	Not Assigned	0	0.0%
TOTAL	1,435	100%	TOTAL	2,559	100%	TOTAL	310	100%

Table 3. Number and percentage of news outlets by languages in the three major providers. First ten languages.

Figure 4 and Table 4 show the distribution of blogs by language. Figure 4 ranks the languages by the percentage of blogs in Altmetric.com. Table 4 shows the first ten languages in each provider by number of blogs. Unlike news, the distribution of blogs by language in the three providers is broader. Thus, for example, the proportion of English-speaking blogs declines for news and is now 89.3% of blogs in PlumX and 85.9% in CED. Only Altmetric.com includes more English-speaking blogs (84.7%) than news (76.7%). Spanish-speaking blogs are second in PlumX (1.8%) and CED (3.5%), while in Altmetric.com, German-speaking blogs are second (2.8%).
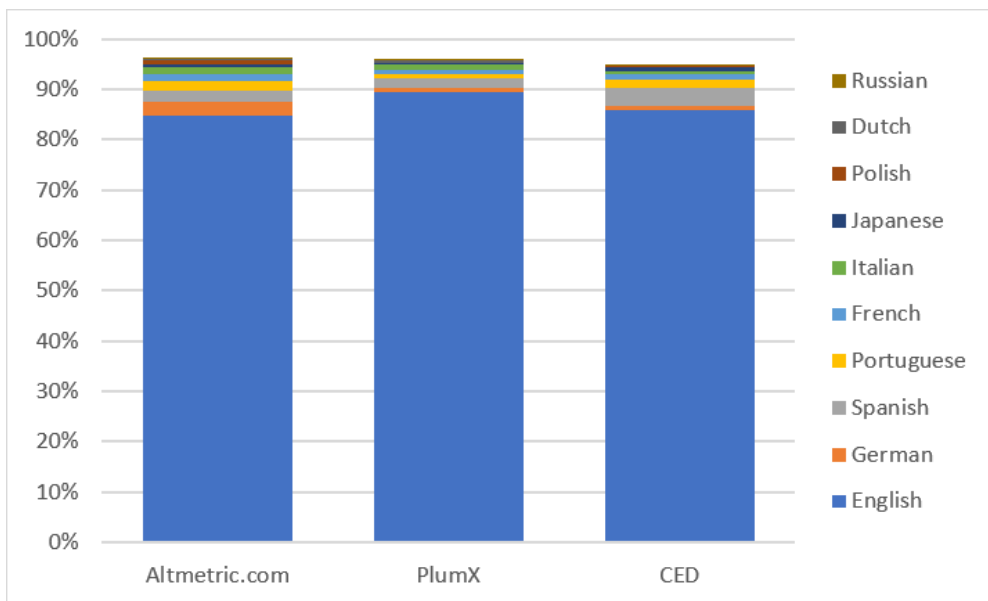

Figure 4. Percentage of blogs by language in the three major providers limited to the first ten.

Altmetric.com			PlumX			CED		
Language	Sources	%	Language	Sources	%	Language	Sources	%
English	2186	84.7%	English	776	89.3%	English	875	85.9%
German	73	2.8%	Spanish	16	1.8%	Spanish	36	3.5%
Spanish	57	2.2%	Italian	9	1.0%	Portuguese	17	1.7%
Portuguese	47	1.8%	German	9	1.0%	Chinese	15	1.5%
French	38	1.5%	French	8	0.9%	French	10	1.0%
Italian	32	1.2%	Portuguese	6	0.7%	German	9	0.9%
Japanese	20	0.8%	Japanese	4	0.5%	Japanese	9	0.9%
Polish	17	0.7%	Swedish	3	0.3%	Italian	6	0.6%
Dutch	6	0.2%	Dutch	3	0.3%	Swedish	4	0.4%

Russian	5	0.2%	Croatian	3	0.3%	Indonesian	4	0.4%
Not Assigned	81	5.6%	Not Assigned	27	1.1%	Not Assigned	22	7.1%
TOTAL	2582	100%	TOTAL	869	100%	TOTAL	1019	100%

Table 4. Number and percentage of blogs by languages in the three major providers. First ten languages.

Altmetric.com has the lowest concentration of languages in the case of news (95.5%), followed by PlumX (98.3) and CED (100%). Altmetric.com is therefore a more global tool because it includes a greater variety of languages. This pattern is repeated in blogs, with 96.1% for Altmetric.com and 96.7% for CED.

Subject

The last criterion for comparing blogs and news sources is content. This element is fundamental because the thematic distribution of sources influences the disciplinary impact of research papers, resulting in articles from one discipline (i.e., Oncology) being more likely to be mentioned by specialist media in relation to similar topics (i.e., Health Sciences). More than 90% of the sources in each provider were classified satisfactorily.

Figure 5 and Table 5 show the thematic distribution of news. Figure 5 ranks the percentage of blogs in Altmetric.com according to ASJC Subjects. Table 5 shows the first ten ASJC Fields in each provider. The thematic distribution of sources is much more irregular than the distribution of countries and languages. In terms of scope, Altmetric.com includes more General-interest media (28.8%), while PlumX specializes in Local-interest news outlets (43.6%). In this sense, Altmetric (55.6%) and PlumX (55.3%) present a similar proportion of non-specialist media, whereas CED has a lower amount of General (22.9%) and Local-interest (5.9%) media.
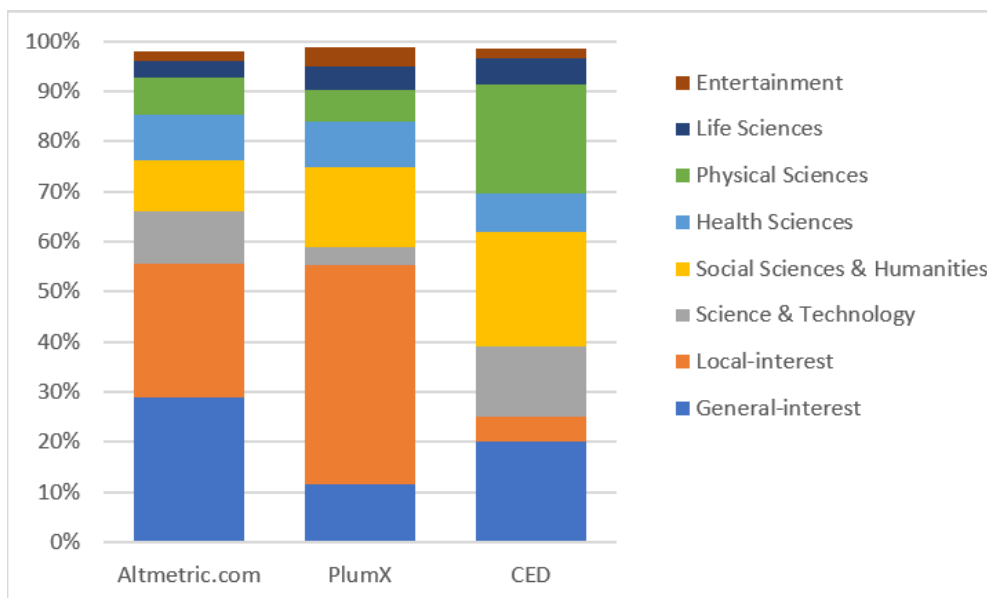

Figure 5. Percentage of news outlets by ASJC Subjects in the three major providers.

Altmetric.com			**PlumX**			**CED**		
Subject	Sources	%	Subject	Sources	%	Subject	Sources	%
General-interest	414	28.8%	Local-interest	1115	43.6%	General-interest	62	22.9%
Local-interest	386	26.8%	General-interest	299	11.7%	Science and Technology	44	16.2%

11

Science and Technology	150	10.4%	Medicine	160	6.3%	Medicine	18	6.6%
Medicine	106	7.4%	Business, Management and Accounting	136	5.3%	Sociology and Political Science	17	6.3%
Business, Management and Accounting	45	3.1%	Science and Technology	95	3.7%	Environmental Science	16	5.9%
Economics, Econometrics and Finance	27	1.9%	Entertainment	92	3.6%	Local-interest	16	5.9%
Entertainment	26	1.8%	Finance	75	2.9%	Physics and Astronomy	14	5.2%
Sociology and Political Science	23	1.6%	Pharmacology, Toxicology and Pharmaceutics	50	2.0%	Psychology	12	4.4%
Physics and Astronomy	19	1.3%	Sociology and Political Science	45	1.8%	Earth and Planetary Sciences	11	4.1%
Finance	18	1.3%	Economics, Econometrics and Finance	42	1.6%	Economics, Econometrics and Finance	11	4.1%
Not Assigned	30	2.1%	Not Assigned	30	1.2%	Not Assigned	2	0.6%
TOTAL	1,439	100%	TOTAL	2,559	100%	TOTAL	310	100%

Table 5. Number and percentage of news outlets by ASJC Fields in the three major providers. First ten languages.

As for the distribution of specialist media, Science and Technology (10.4%), Medicine (7.4%), and Business, Management, and Accounting (3.1%) are the main categories in Altmetric.com. These same categories recur in PlumX, but in different proportions. Thus, Medicine (6.3%) is now the principal group, followed by Business, Management, and Accounting (5.3%), and Science and Technology (3.7%). These proportions change in CED, where Science and Technology (16.2%) is the most common category, followed by Medicine (6.6%) and Sociology and Political Science (6.3%). Overall, and according to the main ASJC Subjects, Altmetric.com covers mainly Social Sciences and Humanities (10.2%), Health Sciences (9%), and Physical Sciences (7.5%). PlumX describes a similar distribution with 16% for Social Sciences and Humanities, 9.1% for Health Sciences, and 6.5% for Physical Sciences. However, CED presents 22.9% of Social Sciences and Humanities media and a strong increase in Physical Sciences (21.6%) to the detriment of Health Sciences (7.7%).

Figure 6 and Table 6 display the thematic distribution of blogs. Table 6 shows the first ten ASJC Fields in each provider. Figure 6 depicts the same information but the fields are grouped in Subjects and the percentage is ranked according to Altmetric.com. The distribution of blogs is rather different to news, and the presence of General and Local-interest media is marginal and, in many cases, due to a misclassification of news as blogs (Ortega, 2019). Most of the blogs covered by Altmetric.com are about Science and Technology (13.1%), followed by Medicine (9.9%), and Agricultural and Biological Sciences (5.9%). PlumX describes a similar proportion, where Science and Technology covers 11.6%, Medicine 11.5%, and Sociology and Political

Science 9.8%. These distributions differ slightly in CED, with 9.9% for Medicine, 7.1% for Science and Technology, and 5.1% for Psychology.
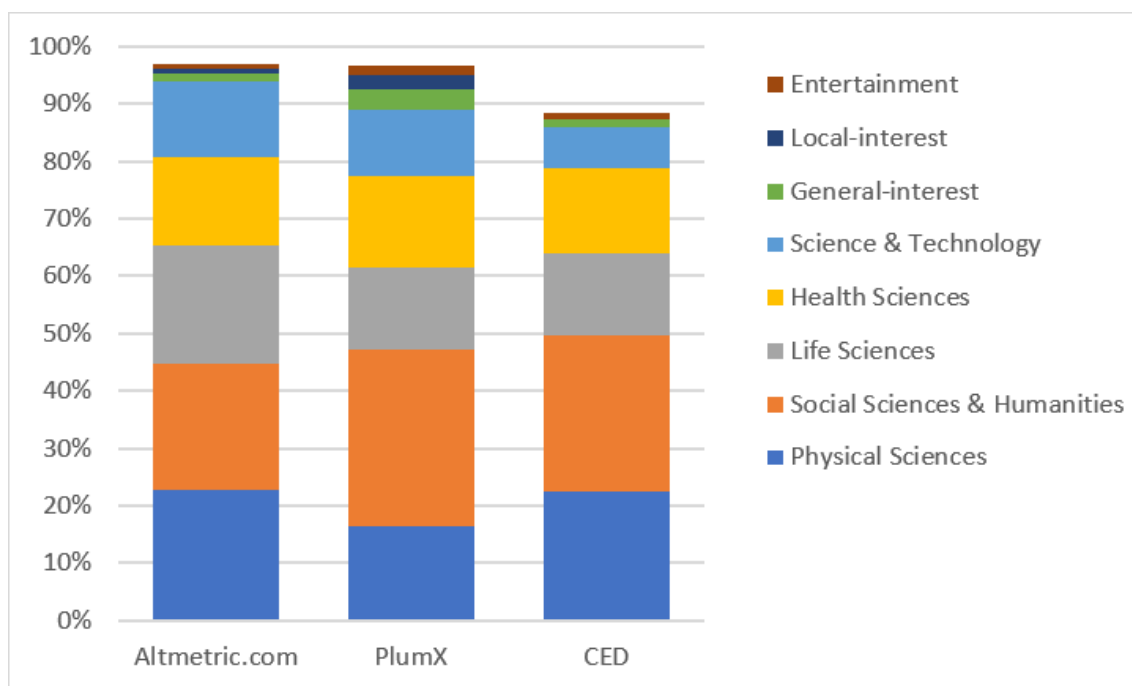

Figure 6. Percentage of blogs by ASJC Subjects in the three major providers.

Altmetric.com			PlumX			CED		
Subject	Sources	%	Subject	Sources	%	Subject	Sources	%
Science and Technology	339	13.1%	Science and Technology	101	11.6%	Medicine	101	9.9%
Medicine	256	9.9%	Medicine	100	11.5%	Science and Technology	72	7.1%
Agricultural and Biological Sciences	152	5.9%	Sociology and Political Science	85	9.8%	Psychology	52	5.1%
Physics and Astronomy	121	4.7%	Environmental Science	37	4.3%	Education	50	4.9%
Neuroscience	113	4.4%	Neuroscience	35	4.0%	Environmental Science	39	3.8%
Environmental Science	94	3.6%	General-interest	31	3.6%	Neuroscience	28	2.7%
Psychology	91	3.5%	Economics, Econometrics and Finance	31	3.6%	Library and Information Sciences	28	2.7%
Library and Information Sciences	91	3.5%	Agricultural and Biological Sciences	24	2.8%	Chemistry	28	2.7%
Genetics	89	3.4%	Genetics	24	2.8%	Physics and Astronomy	26	2.6%
Economics, Econometrics and Finance	87	3.4%	Physics and Astronomy	23	2.6%	Sociology and Political Science	24	2.4%
Not Assigned	76	2.9%	Not Assigned	23	2.6%	Not Assigned	22	2.2%

TOTAL	2,582	100%	TOTAL		869	100%	TOTAL	1,019	100%

Table 6. Number and percentage of blogs by ASJC Fields in the three major providers. First ten languages.

With grouping by Subjects, Altmetric.com mainly indexes Physical Sciences (22.6%), Social Sciences and Humanities (22.1%), and Life Sciences (20.6%) blogs. In the case of PlumX, Social Sciences and Humanities (30.8%) is the most common class, followed by Physical Sciences (16.3%) and Health Sciences (16%). Finally, CED reproduces the same distribution as PlumX, where Social Sciences and Humanities (27.4%) is again the largest group, followed by Physical Sciences (22.4%) and Health Sciences (14.7%).

Discussion

Analysis of the sources that feed each altmetrics provider has enabled the comparison of the coverage of these sites, according to three key elements: country, language coverage, and thematic distribution. The number of sources captured in each provider (Altmetric.com blogs = 2,582, news = 1,439; PlumX blogs = 869, news = 2,559) is rather lower than the official figures (14,000 blogs and 2,900 news outlets for Altmetric.com, and 10,000 blogs and 55,000 news sources for PlumX) (Altmetric.com, 2019; Allen, 2017). This discrepancy demonstrates that this type of mention is scattered over a large number of sources, and very likely most of them only capture a few events. The results show that even a significant sample like this one is not enough to gather all the sources that feed these platforms. Despite this limitation, the percentages obtained for countries, languages, and subjects are comparable to other sources. In fact, if we observe source share by country in Newsflo (current providers of news for PlumX) (Elsevier, 2018), country rank coincides with these results, except for the last position (Switzerland, instead of Ireland).

The most important result is the strong presence of English-speaking countries and English language in the sources. In all three providers, English-speaking countries make up more than 67% in news and blogs. This bias is larger according to Language, in which English-speaking sources amount to more than 90% in the three sites. This result was already observed by Fraumann et al. (2015), who detected 80% of English-speaking sources when they studied the coverage of blogs and news by Altmetric.com. This overwhelming presence could enhance the mention of scholarly results written in English or from English-speaking countries over other papers. Alperin's (2015) results about the mention of SciELO publications (most of them in Portuguese and Spanish) in blogs and news from Altmetric.com (15 mentions to 389,795 documents) highlighted the negligible impact of non-English language articles. Similar results were found by Togia et al. (2017), who studied the altmetrics impact of a Greek university and detected that only 6.6% of the articles were mentioned in blogs and news. Another explanation of this over-representation could be because English is the current scientific *lingua franca* in which many blogs and news outlets are written, despite their being located in a non-English-speaking country. In Altmetric.com, for example, 84.7% of the blogs are written in English, but only 77.4% of them come from English-speaking countries. Similar percentages are seen for PlumX and CED. However, it is true that the poor coverage of local non-English-speaking media could undermine the impact of studies focusing on regional issues (i.e., local history, native species, national policies, etc.).

A comparison of the three providers reveals that PlumX has a better coverage of English-speaking countries, because 86.7% of the news and 80.5% of the blogs came from these places. Conversely, Altmetric.com covers more news (32.9%) and blogs (22.6%) from non-English-

speaking countries, making it the most heterogeneous and diverse platform. CED, on the contrary, has the largest proportion of news from English-speaking countries (87.7%) and a serious proportion (20.2%) of empty or inactive blogs that could not be classified.

According to language, practically the entire list of news media covered by CED is English-speaking (97.7%), while Altmetric.com is once again the platform with less English-speaking news media (76.7%), which confirms it as the most diverse platform. PlumX is also a service with a deep bias in favor of English, being the first to cover English-speaking blogs (89.3%) and the second in news (94.6%).

Another important result is subject distribution, where significant differences between blogs and news were detected. News are mainly non-specialist media that cover general topics: over 50% in Altmetric.com and PlumX. In this regard, PlumX gathers more Local-interest media (43.6%), while Altmetric.com features General-interest news outlets (28.8%). Blogs, however, are specialized sources (more than 90%) that deal with specific topics. The strong presence of Social Sciences and Humanities is interesting to note, being the main subject in PlumX (30.8%) and CED (27.8%), and the second in Altmetric.com (22.1%). This result would explain why social sciences disciplines are mentioned more often in blogs than cited in scholarly literature (Costas et al., 2015a, 2015b) and addresses the importance of the Social Sciences in public opinion.

In view of the results, it is hard to recommend only one service for altmetrics studies. Altmetric.com tracks more sources and those that are more globally distributed. However, PlumX has a better coverage of local news outlets, while CED, though well-balanced, only captures a small fraction of events. Therefore, for any altmetrics study, different sources should be selected and complemented and one or other service chosen in each particular case. Consequently, more studies in this line would help set more consistent criteria for the selection of data providers.

Conclusions

The results of the source coverage of altmetrics providers lead us to conclude that the three services have a considerable presence of blogs and news media from the United States (> 45%). Added to other English-speaking countries, such as the United Kingdom, Australia, and Canada, the percentage increases to more than 67% of blogs and news in all three providers. Altmetric.com is the least biased site in the coverage of news outlets (67.1%) and blogs (77.4%).

Distribution by language follows an even more skewed pattern with percentages higher than 75% for English-speaking media. In this sense, Altmetric.com is again the most balanced platform with 76.7% of English-speaking news outlets and 84.7% of blogs. These geographical and language biases could underestimate the altmetrics impact of publications in other languages or from non-English-speaking countries.

Finally, blogs and news follow different patterns according to disciplinary classification. News is commonly General-interest media (> 50% in PlumX and Altmetric.com), PlumX covers more Local-interest media (43.6%), while Altmetric.com specializes in General-interest news (28.8%). According to blogs, most are specialized venues, with a high proportion of Social Sciences and Humanities sites (> 20%).

References

Adie, E. & Roe, W. (2013). Altmetric: enriching scholarly content with article-level discussion and metrics. Learned Publishing, 26(1):11-17.

Allen, P. (2017). World-Wide News Coverage in PlumX. Retrieved May 20, 2019 from: https://plumanalytics.com/world-wide-news-coverage-plumx/

Alperin, J. P. (2015). Geographic variation in social media metrics: An analysis of Latin American journal articles. Aslib Journal of Information Management, 67(3): 289-304.

Altmetric.com (2019a). How it works. Retrieved May 20, 2019 from: https://www.altmetric.com/about-our-data/how-it-works/

Altmetric.com (2019b). What outputs and sources does Altmetric track? Altmetric.com Support. Retrieved May 20, 2019 from: https://help.altmetric.com/support/solutions/articles/6000060968-what-outputs-and-sources-does-altmetric-track-

Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. Journal of Informetrics, 8(4): 935-950.

Champieux, R. (2015). PlumX. Journal of the Medical Library Association: JMLA, 103(1): 63.

Costas, R., Zahedi, Z., & Wouters, P. (2015a). The thematic orientation of publications mentioned on social media: large-scale disciplinary comparison of social media metrics with citations. Aslib Journal of Information Management, 67(3): 260-288.

Costas, R., Zahedi, Z., & Wouters, P. (2015b). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. Journal of the Association for Information Science and Technology, 66(10): 2003-2019.

Crossref (2019). Crossref Event Data User Guide. Retrieved May 20, 2019 from: https://www.eventdata.crossref.org/guide/

Elsevier. (2017). Elsevier Acquires Leading 'Altmetrics' Provider Plum Analytics. Retrieved January 2, 2019 from: https://www.elsevier.com/about/press-releases/corporate/elsevier-acquires-leading-altmetrics-provider-plum-analytics

Elsevier. (2018). Newsflo > Coverage and sources. Retrieved May 20, 2019 from: https://www.elsevier.com/solutions/newsflo/sources

Fraumann, G., Zahedi, Z. & Costas, R. (2015). What do we know about Altmetric.com sources? A study of the top 200 blogs and news sites mentioning scholarly output. Altmetrics workshop, 9 October 2015, Amsterdam Science Park, Amsterdam.

Hassan, S. U., Imran, M., Gillani, U., Aljohani, N. R., Bowman, T. D., & Didegah, F. (2017). Measuring social media activity of scientific literature: an exhaustive comparison of scopus and novel altmetrics big data. Scientometrics, 113(2): 1037-1057.

Holmberg, K. J. (2015). *Altmetrics for information professionals: Past, present and future*. Chandos Publishing.

Jobmann, A., Hoffmann, C.P., Künne, S., Peters, I., Schmitz, J. &Wollnik-Korn, G. (2014). Altmetrics for large, multidisciplinary research groups: Comparison of current tools.

Bibliometrie –Praxis und Forschung, 3, http://www.bibliometrie-pf.de/article/viewFile/205/258

Kraker, P., Lex, E., Gorraiz, J., Gumpenberger, C., & Peters, I. (2015). Research data explored II: The anatomy and reception of figshare. *arXiv preprint arXiv:1503.01298*.

Meschede, C., & Siebenlist, T. (2018). Cross-metric compatability and inconsistencies of altmetrics. Scientometrics, 115(1): 283-297.

Ortega, J. L. (2018), Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data, Scientometrics,116(3): 2123-2138.

Ortega, J. L. (2019a), The coverage of blogs and news in the three major altmetric data providers, *17th International Conference of the International Society for Scientometrics and Informetrics, Rome, Italy* https://osf.io/pd27h/

Ortega, J. L. (2019b). Availability and Audit of Links in Altmetric Data Providers: Link Checking of Blogs and News in Altmetric.com, Crossref Event Data and PlumX. *Journal of Altmetrics*, 2(1): 4

Parkhill, M. (2015). PlumX Expands "Mentions" Category by Tracking Blogs. Plum Analytics. Retrieved May 20, 2019 from: https://plumanalytics.com/plumx-expands-mentions-category-by-tracking-blogs/

Parkhill, M. (2016). Research Impact Tracking from Scholarly Blog Posts Expands with Partnership Between Plum™ Analytics and ACI Information Group. Plum Analytics. Retrieved May 20, 2019 from: https://plumanalytics.com/press/research-impact-tracking-scholarly-blog-posts-expands-partnership-plum-analytics-aci-information-group/

Peters, I., Kraker, P., Lex, E., Gumpenberger, C. & Gorraiz, J. (2016). Research data explored: an extended analysis of citations and altmetrics. Scientometrics, 107(2):723-744.

Plum Analytics (2019). Coverage: Expanding the World of Altmetrics. Retrieved May 20, 2019 from: https://plumanalytics.com/learn/about-metrics/coverage/

Robinson-García, N., Torres-Salinas, D., Zahedi, Z. & Costas, R. (2014). Nuevos datos, nuevas posibilidades: Revelando el interior de Altmetric.com. El Profesional de la Información, 23(4):359-366.

Scopus (2019). What is the complete list of Scopus Subject Areas and All Science Journal Classification Codes (ASJC)? - Scopus: Access and use Support Center. Retrieved May 20, 2019 from: https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. Journal of the Association for information science and technology, 65(5): 1018-1027.

Togia, A., Koseoglou, E., & Zapounidou, S. (2017). Alternative Metrics for the Evaluation of Scholarly Activities: An Analysis of Articles Authored by Greek Researchers. *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, 138.

Torres-Salinas, D., Gumpenberger, C. & Gorraiz, J. (2017). PlumX as a potential tool to assess the macroscopic multidimensional impact of books. Frontiers in research metrics and analytics, 2(5).

Torres-Salinas, D., Gorraiz, J. & Robinson-Garcia, N. (2018). The insoluble problems of books: What does Altmetric.com have to offer?. Aslib Journal of Information Management, 70(6): 691-707.

Williams, C. (2015) Altmetric to provide expanded mainstream media tracking. Altmetric.com blog. Retrieved May 20, 2019 from: https://www.altmetric.com/blog/moreover/

Zahedi, Z., Fenner, M. & Costas, R. (2015). Consistency among altmetrics data provider/aggregators: What are the challenges?. The 2015 Altmetrics Workshop, Amsterdam.

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. PloS one, 13(5): e0197326.